

# Biology, Mathematics, and a Mathematical Biology Laboratory

## 1.1 The Natural Linkage Between Mathematics and Biology

*Mathematics and biology have a synergistic relationship. Biology produces interesting problems, mathematics provides models to understand them, and biology returns to test the mathematical models. Recent advances in computer algebra systems have facilitated the manipulation of complicated mathematical systems. This has made it possible for scientists to focus on understanding mathematical biology, rather than on the formalities of obtaining solutions to equations.*

*What is the function of mathematical biology?*

Our answer to this question, and the guiding philosophy of this book, is simple: The function of mathematical biology is to exploit the natural relationship between biology and mathematics. The linkage between the two sciences is embodied in these reciprocal contributions that they make to each other: Biology generates complex problems and mathematics can provide ways to understand them. In turn, mathematical models suggest new lines of inquiry that can only be tested on real biological systems.

We believe that an understanding of the relationship between two subjects must be preceded by a thorough understanding of the subjects themselves. Indeed, the excitement of mathematical biology begins with the discovery of an interesting and uniquely biological problem. The excitement grows when we realize that mathematical tools at our disposal can profitably be applied to the problem. The interplay between mathematical tools and biological problems constitutes mathematical biology.

*The time is right for integrating mathematics and biology.*

Biology is a rapidly expanding science; research advances in the life sciences leave virtually no aspects of our public and private lives untouched. Newspapers bombard us with information about *in vitro* fertilization, bioengineering, DNA testing, genetic manipulation, environmental degradation, AIDS, and forensics.

Quite separately from the news pouring onto us from the outside world, we have an innate interest in biology. We have a natural curiosity about ourselves. Every day we ask ourselves a nonstop series of questions: What happens to our bodies as we get older? Where does our food go? How do poisons work? Why do I look like my mother? What does it mean to “think”? Why are HIV infections spreading so rapidly in certain population groups?

Professional biologists have traditionally made their livings by trying to answer these kinds of questions. But scientists with other kinds of training have also seen ways that they could enter the fray. As a result, chemists, physicists, engineers, and mathematicians have all made important contributions to the life sciences. These contributions often have been of a sort that required specialized training or a novel insight that only specialized training could generate.

In this book we present some mathematical approaches to understanding biological systems. This approach has the hazard that an in-depth analysis could quickly lead to unmanageably complex numerical and symbolic calculations. However, technical advances in the computer hardware and software industries have put powerful computational tools into the hands of anyone who is interested. Computer algebra systems allow scientists to bypass some of the details of solving mathematical problems. This then allows them to spend more time on the interpretation of biological phenomena, as revealed by the mathematical analysis.<sup>1</sup>

## 1.2 The Use of Models in Biology

*Scientists must represent real systems by models. Real systems are too complicated, and besides, observation may change the real system. A good model should be simple and it should exhibit the behaviors of the real system that interest us. Further, it should suggest experimental tests of itself that are so revealing that we must eventually discard the model in favor of a better one. We therefore measure scientific progress by the production of better and better models, not by whether we find some absolute truth.*

*A model is a representation of a real system.*

The driving force behind the creation of models is this admission: Truth is elusive, but we can gradually approximate it by creating better and better representations.

There are at least two reasons why the truth is so elusive in real systems. The first reason is obvious: The universe is extremely complicated. People have tried unsuccessfully to understand it for millennia, running up countless blind alleys and only occasionally finding enlightenment. Claims of great success abound, usually followed by their demise. Physicists in the late nineteenth century advised their students that Maxwell’s equations had summed up everything important about physics, and that further research was useless. Einstein then developed the theory of general

---

<sup>1</sup> References [1]–[4] at the end of this chapter are some articles that describe the importance of mathematical biology.

relativity, which contained Maxwell's equations as a mere subcategory. The unified field theory ("The Theory of Everything") will contain Einstein's theory as a subcategory. Where will it end?

The second reason for the elusivity of the truth is a bit more complicated: It is that we tend to change reality when we examine any system too closely. This concept, which originates in quantum mechanics, suggests that the disturbances that inevitably accompany all observations will change the thing being observed. Thus "truth" will be changed by the very act of looking for it.<sup>2</sup> At the energy scale of atoms and molecules the disturbances induced by the observer are especially severe. This has the effect of rendering it impossible to observe a single such particle without completely changing some of the particle's fundamental properties. There are macroscopic analogues to this effect. For example, what is the "true" color of the paper in this book? The answer depends on the color of the light used to illuminate the paper, white light being merely a convenience; most other colors would also do. Thus you could be said to have chosen the color of the paper by your choice of observation method.

Do these considerations make a search for ultimate explanations hopeless? The answer is, "No, because what is really important is the progress of the search, rather than some ultimate explanation that is probably unattainable anyway."

*Science is a rational, continuing search for better models.*

Once we accept the facts that a perfect understanding of very complex systems is out of reach and that the notion of "ultimate explanations" is merely a dream, we will have freed ourselves to make scientific progress. We are then able to take a reductionist approach, fragmenting big systems into small ones that are individually amenable to understanding. When enough small parts are understood, we can take a holistic approach, trying to understand the relationships among the parts, thus reassembling the entire system.

In this book we reduce complicated biological systems to relatively simple mathematical models, usually of one to several equations. We then solve the equations for variables of interest and ask whether the functional dependencies of those variables predict salient features of the real system.

There are several things we expect from a good model of a real system:

- (a) It must exhibit properties that are similar to those of the real system, and those properties must be the ones in which we are interested.<sup>3</sup> A six-inch replica of a 747 airliner, after adjusting for Reynolds' number, may have the exact fluid-dynamical properties of the real plane, but would be useless in determining the comfort of the seats of a real 747.

<sup>2</sup> This situation is demonstrated by the following exchange: *Question:* How would you decide which of two gemstones is a real ruby and which is a cheap imitation? *Answer:* Tap each sharply with a hammer. The one that shatters used to be the real ruby.

<sup>3</sup> One characteristic of the real system that we definitely do *not* want is its response to the observation process, described earlier. In keeping with the concept of a model as an idealization, we want the model to represent the real system in a "native state," divorced from the observer.

- (b) It must self-destruct. A good model must suggest tests of itself and predict their outcomes. Eventually a good model will suggest a very clever experiment whose outcome will not be what the model predicted. The model must then be discarded in favor of a new one.

The search for better and better models thus involves the continual testing and replacement of existing models. This search must have a rational foundation, being based on phenomena that can be directly observed. A model that cannot be tested by the direct collection of data, and which therefore must be accepted on the basis of faith, has no place in science.

*Many kinds of models are important in understanding biological phenomena.*

Models are especially useful in biology. The most immediate reason is that living systems are much too complicated to be truly understood as whole entities. Thus to design a useful model, we must strip away irrelevant, confounding behaviors, leaving only those that directly interest us. We must walk a fine line here: In our zeal to simplify, we may strip away important features of the living system, and at the other extreme, a too-complicated model is intractable and useless.

Models in biology span a wide spectrum of types. Here are some that are commonly used:

Model	What the model represents
$aa \times Aa$	Gene behavior in a genetic cross.
$\frac{dA}{dt} = -kA$	Rate of elimination of a drug from the blood.
$\boxed{R} \rightarrow \boxed{C} \rightarrow \boxed{E}$	Reflex arc involving a stimulus <u>R</u> ceptor, the <u>C</u> entral nervous system, and an <u>E</u> ffector muscle.
a camera	The eye of a vertebrate or of an octopus.

*Why is there so much biological information in this book?*

It is possible to write a mathematical biology book that contains only a page or two of biological information at the beginning of each chapter. We see that format as the source of two problems: First, it is intellectually limiting. A student cannot apply the powerful tools of mathematics to biological problems he or she does not understand. This limitation can be removed by a thorough discussion of the underlying biological systems, which can suggest further applications of mathematics. Thus a strong grounding in biology helps students to move further into mathematical biology.

Second, giving short shrift to biology reinforces the misconception that each of the various sciences sits in a vacuum. In fact, it has been our experience that many students of mathematics, physics, and engineering have a genuine interest in biology, but little opportunity to study it. Taking our biological discussions well beyond the barest facts can help these students to understand the richness of biology, and thereby encourage interdisciplinary thinking.

## 1.3 What Can Be Derived from a Model and How Is It Analyzed?

*A model is more than the sum of its parts. Its success lies in its ability to discover new results, results that transcend the individual facts built into it. One result of a model can be the observation that seemingly dissimilar processes are in fact related. In an abstract form, the mathematical equations of the process might be identical to those of other phenomena. In this case the two disciplines reinforce: A conclusion difficult to see in one might be an easy consequence in the other.*

To analyze the mathematical equations that arise, we draw on the fundamentals of matrix calculations, counting principles for permutations and combinations, the calculus, and fundamentals of differential equations. However, we will make extensive use of the power of numerical and symbolic computational software—a computer algebra system. The calculations and graphs in this text are done using such software.

Syntax for both MAPLE and MATLAB accompanies the mathematical derivations in the text. This code should be treated something like a displayed equation. Like an equation, code is precise and technical. On first reading, it is often best to work through a line of reasoning, with only a glance at any included code, to understand the points being made. Then a critical examination of an equation or piece of code will make more sense, having the benefit of context and intended goal. The computer algebra syntax is displayed and set off in a distinctive font in order for the reader to be able to quickly find its beginning and ending. Where possible, equivalent syntax for MAPLE and MATLAB are presented together in tandem. It should be noted that the basic MATLAB system is numerical and does not perform symbolic computations. Thus equivalent MATLAB code is omitted in this case. An accessory package is available for MATLAB that can perform symbolic manipulation. And conveniently, this package is created by the same people who created MAPLE.

*Deriving consequences: The other side of modeling.*

Once a model has been formulated and the mathematical problems defined, then they must be solved. In this symbolic form, the problem takes on a life of its own, no longer necessarily tied to its physical origins. In symbolic form, the system may even apply to other, totally unexpected, phenomena. What do the seven bridges at Königsberg have to do with discoveries about DNA? The mathematician Euler formed an abstract model of the bridges and their adjoining land masses and founded the principles of Eulerian graphs on this model. Today, Eulerian graphs are used, among other ways, to investigate the ancestry of living things by calculating the probability of matches of DNA base pair sequences (see Kandel [5]). We take up the subject of phylogeny in Chapter 15. The differential equations describing spring–mass systems and engineering vibrations are identical to those governing electrical circuits with capacitors, inductors, and resistors. And again these very same equations pertain to the interplay between glucose and insulin in humans. The abstract and symbolic treatment of these systems through mathematics allows the transfer of intuition between them. Through

mathematics, discoveries in any one of these areas can lead to a breakthrough in the others. But mathematics and applications are mutually reinforcing: The abstraction can uncover truths about the application, suggesting questions to ask and experiments to try; the application can foster mathematical intuition and form the basis of the results from which mathematical theorems are distilled.

In symbolic form, a biological problem is amenable to powerful mathematical processing techniques, such as differentiation or integration, and is governed by mathematical assertions known as theorems. Theorems furnish the conclusions that may be drawn about a model so long as their hypotheses are fulfilled. Assumptions built into a model are there to allow its equations to be posed and its conclusions to be mathematically sound. The validity of a model is closely associated with its assumptions, but experimentation is the final arbiter of its worth. The assumption underlying the exponential growth model, namely,  $\frac{dy}{dt} = ky$  (see Section 2.4 and Chapter 3), is unlikely to be precisely fulfilled in any case, yet exponential growth is widely observed for biological populations. However, exponential growth ultimately predicts unlimited population size, which never materializes precisely due to a breakdown in the modeling assumption. A model is robust if it is widely applicable. In every case, the assumptions of a model *must* be spelled out and thoroughly understood. The validity of a model's conclusions must be experimentally confirmed. Limits of applicability, robustness, and regions of failure need to be determined by carefully designed experiments.

Some biological systems involve only a small number of entities or are greatly influenced by a few of them, maybe even one. Consider the possible DNA sequences 100 base pairs long. Among the possibilities, one or two base pairs might be critical to life. (It is known that tRNA molecules can have as few as 73 nucleotide residues (Lehninger [6]).) Or consider the survival prospects of a clutch of Canadian geese blown off migratory course to the Hawaiian islands. Their survival analysis must keep track of detailed events for each goose and possibly even details of their individual genetic makeups, for the loss of a single goose or the birth of defective goslings could spell extinction for the small colony. (The nene, indigenous to Hawaii, is thought to be related to the Canadian geese.) This is the mathematics of discrete systems, i.e., the mathematics of a finite number of states. The main tools we will need here are knowledge of matrices and their arithmetic, counting principles for permutations and combinations, and some basics of probability calculations.

Other biological systems or processes involve thousands, even millions, of entities, and the fate of a few of them has little influence on the entire system. Examples are the diffusion process of oxygen molecules or the reproduction of a bacterial colony. In these systems, individual analysis gives way to group averages. An average survival rate of 25% among goslings of a large flock of Canadian geese still ensures exponential growth of the flock in the absence of other effects; but this survival probability sustained by exactly four offspring of an isolated clutch might not result in exponential growth at all but rather total loss instead. When there are large numbers involved, the mathematics of the continuum may be brought to bear, principally calculus and differential equations. This greatly simplifies the analysis. The techniques are powerful and mature, and a great many are known.

*Computer algebra systems make the mathematics accessible.*

It is a dilemma: Students in biology and allied fields such as immunology, epidemiology, or pharmacology need to know how to quantify concepts and to make models. Yet, these students typically have only one year of undergraduate study in mathematics. (Hopefully this will change in our postgenomics world.) This one year may be very general and not involve any examples from biology. When the need arises, they are likely to accept the models and results of others, perhaps without deep understanding.

On the other side of campus, students in mathematics read in the popular technical press of biological phenomena, and wish they could see how to use their flair for mathematics to get them into biology. The examples they typically see in mathematics classes have their roots in physics. Applications of mathematics to biology seem far away.

How can this dilemma be resolved? Should the biology students be asked to take a minor in mathematics in order to be ready to use the power of differential equations for modeling? And what of algebraic models, discrete models, probabilistic models, or statistics? Must the mathematics students take a course in botany, and then zoology, before they can make a model for the level to which the small vertebrate population must be immunized in a geographic region in order to reduce the size of the population of ticks carrying Lyme disease? Such a model is suggested by Kantor [7].

There is an alternative. Computer algebra systems create a new paradigm for designing, analyzing, and drawing conclusions from models in science and engineering. The technology in the computer algebra systems allows the concepts to be paramount while computations and details become less important. With such a computational engine it is possible to read about models that are being actively explored in the current literature and do a computer analysis of these new models.

The theorems from which our conclusions are derived often result from carefully tracking evolving system behavior over many iterations in discrete systems or infinite time in continuous ones. Where possible, the mathematical equations are solved and the solutions exhibited. Predictions of the model are made under a range of starting conditions and possibly unusual parameter regimes. These are the bases of “what if” experiments. For example, given a satisfactory model for a fishery, *what if* one imposes various levels of harvesting? To answer this and related questions, the computer algebra system can carry out the technical computations: calculate roots, differentiate symbolically or numerically, integrate and solve differential equations, perform matrix arithmetic, track system evolution, and graphically display results.

In this book we will use computational packages to do the “heavy lifting.” MATLAB is a very powerful system for general numerical computation. In addition, accessory “toolboxes” are available providing the specialized computations used in several disciplines. MAPLE is a system for both numerical and symbolic calculations. MAPLE is quite complete in its mathematical coverage and especially strong in symbolic computations. In addition to these two, recently created software packages are available to perform the computations of the emerging field of algebraic statistics.

These packages are for the most part available free of charge. We will encounter *BLAST* and *SINGULAR* in the genomics sections of the book.

## References

- [1] THE FUTURE OF MATHEMATICAL BIOLOGY:  
*Mathematics and Biology: The Interface, Challenges, and Opportunities*, National Science Foundation workshop, Lawrence Berkeley Laboratory, Berkeley, CA, U.S. Department of Energy, Washington, DC, 1992.
- [2] THE IMPORTANCE OF MATHEMATICS IN THE LIFE SCIENCES:  
 L. J. Gross, Quantitative training for life-science students, *Biosci.*, **44-2** (1994), 59.
- [3] MODELING IN BIOLOGY:  
 W. D. Hillis, Why physicists like models and why biologists should, *Curr. Biol.*, **3-2** (1993), 79–81.
- [4] APPLICATIONS OF MATHEMATICAL BIOLOGY:  
 F. Hoppensteadt, Getting started in mathematical biology, *Not. Amer. Math. Soc.*, **42-9** (1995), 969.
- [5] DNA SEQUENCES:  
 D. Kandel, Y. Matias, R. Unger, and P. Winkler, *Shuffling Biological Sequences*, preprint, AT&T Bell Laboratories, Murray Hill, NJ, 1995.
- [6] BIOCHEMISTRY OF NUCLEIC ACIDS:  
 A. Lehninger, *Biochemistry*, Worth Publishers, New York, 1975, 935.
- [7] CONQUERING LYME DISEASE:  
 F. S. Kantor, Disarming Lyme disease, *Sci. Amer.*, **271** (1994), 34–39.
- [8] THE PROMISE OF GENOMICS:  
 G. Smith, *Genomics Age*, AMACOM, New York, 2005.